CSE591 (Fall 2005): Machine Learning and Applications

General Course Information

- Instructor: Dr. Jieping Ye
- Office: BY 506 (BY 568 since September 1, 2005)
- Email: jieping.ye@asu.edu
- Meeting Times: MW 11:40am–12:55pm
- Office Hours: MW 10:00am–11:30am
- Location: BYAC 270
- Prerequisite: Basics of linear algebra.
- Web: http://www.public.asu.edu/~jye02/CLASSES/Fall-2005/
- Course Textbook: No textbook is required for this course

Catalog Description

Clustering, bi-clustering, classification, semi-supervised learning, feature reduction, manifold learning, kernel learning, bagging, and boosting.

Objective

An in-depth understanding of some of the most important machine learning methods and their applications in bioinformatics and other domains.

Grading

- Homework (2): 30%. There will be two homeworks assigned.
- Project: 50%. Two to three students form a group to carry out a small research project. It can be an implementation and a comparative study of existing methods, a review of a specific topic, or the development of a new idea. Project proposal is due during the fourth week of classes. The intermediate project report is due duing the twelveth week of classes. The final project report is due during the last week of classes.
- Presentation: 20%. Each student will present one paper. A list of papers for presentation will be provided during the first few weeks of classes.

Reference Books

- Bioinformatics: The Machine Learning Approach, Second Edition. Pierre Baldi and Sren Brunak. 2001.
- Data Analysis and Classification for Bioinformatics. Arun Jagota, 2000.
- Guide to Analysis of DNA Microarray Data. 2nd Edition, 2004.
- Kernel Methods for Pattern Analysis. John Shawe-Taylor and Nello Cristianini, 2004.
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction. T. Hastie, R. Tibshirani, and J. Friedman, 2001.
- Pattern Classification. Richard Duda, Peter Hart, and David Stork. 2nd edition, 2000.
- Kernel Methods in Computational Biology. Bernhard Schölkopf, Koji Tsuda, and Jean-Philippe Vert, editors.
- Introduction to Data Mining. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. 2005.

Topics Covered

Week	Topic
1	Introduction, basics of probability, linear algebra and biology
2	Clustering: K-means, Spectral clustering
3	Bi-clustering
4	Classification: Logistic Regression, Linear Discriminant Analysis
5	Classification: Support Vector Machines
6	Semi-supervised learning
7	Feature reduction: Principal Component Analysis, Canonical Correlation Analysis
8	Manifold learning: Isomap, Local Linear Embedding, Laplacian Eigenmaps
9	Kernel learning: Basics
10–11	Kernel learning: Advanced topics
12	Bagging and boosting
13–15	Student presentation